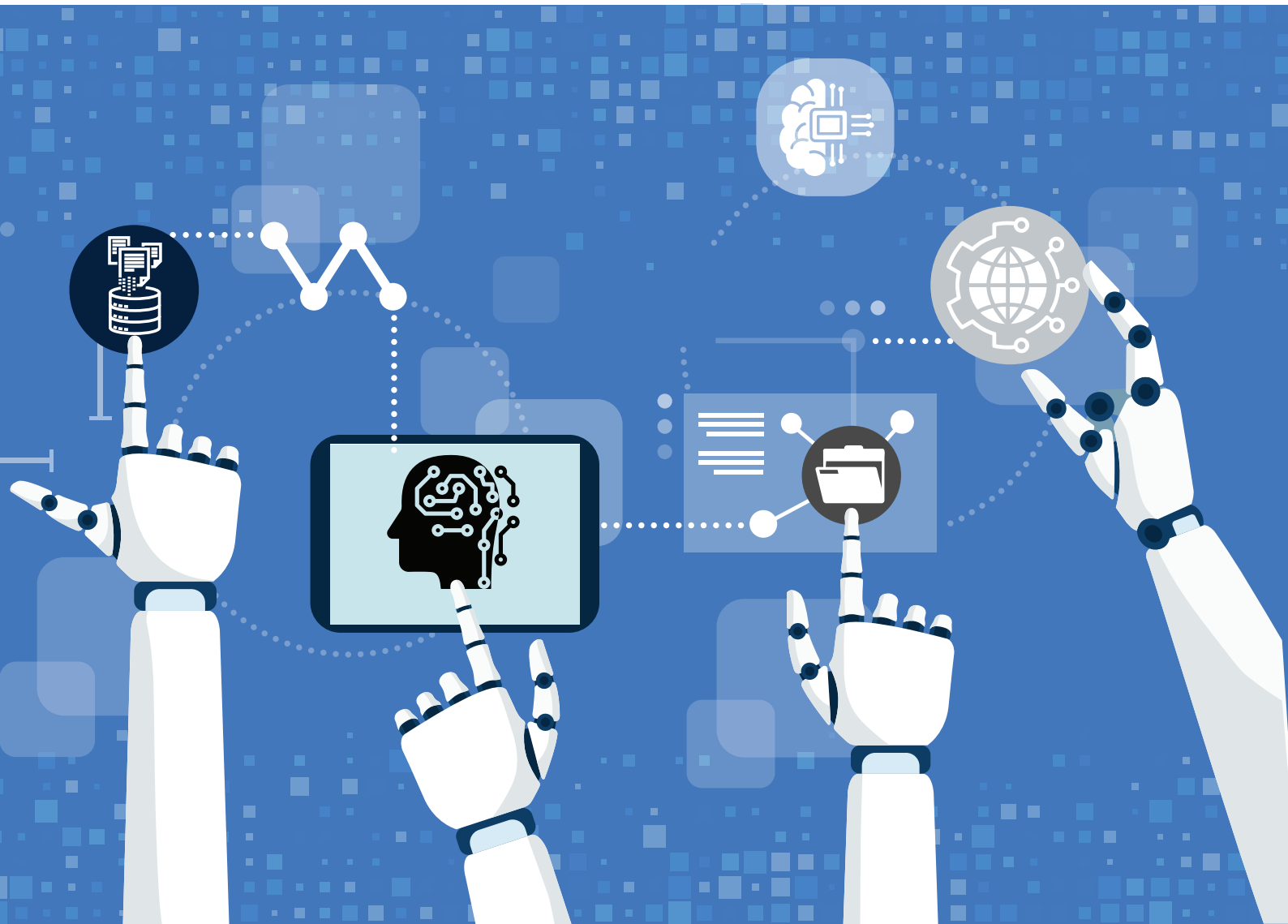
MIT Technology Review Insights

Produced in partnership with

Fivetran

AI is showing executives that a compelling data strategy is the foundation for technological success.

AI readiness for C-suite leaders



Preface

"AI readiness for C-suite leaders" is an MIT Technology Review Insights report sponsored by Fivetran. To produce this report, MIT Technology Review Insights conducted a global survey of C-suite and senior data executives across countries and industries. The report also draws on in-depth interviews conducted with business experts on data and AI.

Paul Kielstra was the author of the report, Teresa Elsey was the editor, and Nicola Crepaldi was the producer. The research is editorially independent, and the views expressed are those of MIT Technology Review Insights.

We would like to thank the following executives for their time and insights:

Stewart Bond, Vice President, Data Intelligence and Integration Software Service, IDC
George Fraser, CEO and Cofounder, Fivetran
Mike Hite, Chief Technology Officer, Saks
Suresh Venkatarayalu, Chief Technology and Innovation Officer, Honeywell

About the survey

The survey forming the basis of this report was conducted by MIT Technology Review Insights in February and March 2024. The survey sample consists of 300 C-suite executives and data and technology leaders. Eleven industries are represented: financial services, manufacturing, IT and telecommunications, consumer goods and retail, pharmaceutical and health care, government, travel and hospitality, professional services, energy and utilities, transport and logistics, and media and marketing.

Nearly all survey respondents (88%) come from the C-suite. These include chief executive officers (22%), chief information officers (23%), chief technology officers (21%), and chief data officers (10%). The respondents' organizations are headquartered in North America (30%); Europe, the Middle East, and Africa (27%); Asia-Pacific (27%); and Central and South America (17%). More than half of the respondents (53%) work at organizations with revenue of $1 billion or more.

CONTENTS

01

Executive summary

64% of C-suite leaders say data integration is a key investment area for AI and generative AI.

Generative AI, like predictive AI before it, has rightly seized the attention of business executives. The technology has the potential to add trillions of dollars to annual global economic activity, and its adoption for business applications is expected to improve the top or bottom lines – or both – at many organizations.

While generative AI offers an impressive and powerful new set of capabilities, its business value is not a given. While some powerful foundational models are open to public use, these do not serve as a differentiator for those looking to get ahead of the competition and unlock AI's full potential. To gain those advantages, organizations must look to enhance AI models with their own data to create unique business insights and opportunities.

Preparing an organization's data for AI, however, unlocks a new set of challenges and opportunities. This MIT Technology Review Insights survey report investigates whether companies' data foundations are ready to garner benefits from generative AI, as well as the challenges of building the necessary data infrastructure for this technology. In doing so, it draws on insights from a survey of 300 C-suite executives and senior technology leaders, as well as on in-depth interviews with four leading experts.

Its key findings include the following:

Data integration is the leading priority for AI readiness. In our survey, 82% of C-suite and other senior executives agree that "scaling AI or generative AI use cases to create business value is a top priority for our organization." The number-one challenge in achieving that AI readiness, survey respondents say, is data integration and pipelines (45%). Asked about challenging aspects of data integration, respondents named four: managing data volume, moving data from on-premises to the cloud, enabling real-time access, and managing changes to data.

Executives are laser-focused on data management challenges – and lasting solutions. Among survey respondents, 83% say that their "organization has identified numerous sources of data that we must bring together in order to enable our AI initiatives." Though data-dependent technologies of recent decades drove data integration and aggregation programs, these were typically tailored to specific use cases. Now, however, companies are looking for something more scalable and use-case agnostic: 82% of respondents are prioritizing solutions "that will continue to work in the future, regardless of other changes to our data strategy and partners."

Data governance and security is a top concern for regulated sectors. Data governance and security concerns are the second most common data readiness challenge (cited by 44% of respondents). Respondents from highly regulated sectors were two to three times more likely to cite data governance and security as a concern, and chief data officers (CDOs) say this is a challenge at twice the rate of their C-suite peers. And our experts agree: Data governance and security should be addressed from the beginning of any AI strategy to ensure data is used and accessed properly.

02

Introduction: Data readiness drives AI value

Today's accelerating rollout of AI in all its forms, spurred by OpenAI's 2022's release of ChatGPT, is expected to reshape much of the economy. The consultancy McKinsey projects that, when fully deployed, predictive AI (including analytics and machine learning) will add between $11.0 and $17.7 trillion to annual global economic activity. Generative AI is projected to contribute an additional $2.6 to $4.4 trillion.[1]

Companies are rushing to grab the opportunity. In our survey, 82% of C-suite and other senior executives agreed that "scaling AI or generative AI use cases to create business value is a top priority for our organization." These results are consistent with numerous other studies pointing to widespread AI adoption by companies across sectors and geographies.[2,3,4]

Suresh Venkatarayalu, chief technology and innovation officer at Honeywell, says that his company has recently deployed generative AI and is already seeing a positive impact on productivity and efficiency.

Amid such widespread adoption, however, merely deploying open-source foundational AI models is unlikely to provide differentiation. Saks chief technology officer Mike Hite explains that "off-the-shelf" AI tools will give all users the same answers. Organizations that are pulling ahead do not necessarily have large AI teams or better large language model (LLM) technology. Instead, he says, "the people who have interesting data sets on a particular thing have opportunity."

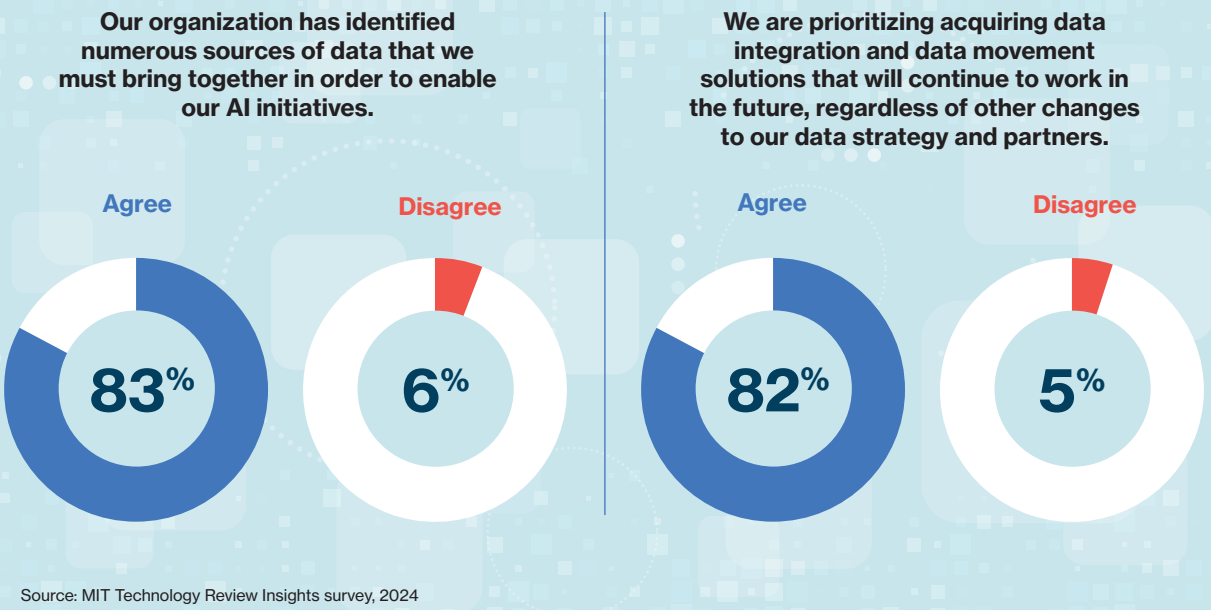
Corralling data effectively

Executives seem to understand the strong link between data and effective AI strategies. CDOs are at the center of the deployment efforts, and almost all CDOs (97%) in our survey describe scaling AI as a top priority.

Organizations don't lack the data they need to meet that priority. Of those surveyed, 83% say that their organization "has identified numerous sources of data

Figure 1. Bringing data together is a top challenge

To what extent do you agree with the following statements?



Source: MIT Technology Review Insights survey, 2024

that we must bring together in order to enable our AI initiatives" (see Figure 1). The challenge is in *how* they bring that data together to satisfy reporting needs, operational workflows, and now their AI ambitions.

The current state at many companies indicates that this task will be far from straightforward. Stewart Bond, vice president of IDC's Data Intelligence and Integration Software Service, says that enterprise information is often very fragmented. A recent IDC survey finds businesses are using "over a dozen different technologies just to harvest all the intelligence about their data and the same number to integrate, transform, and replicate it," says Bond. As a result, organizations see they must invest in getting control of their data.

George Fraser, CEO at Fivetran, notes the importance of investing in data foundations before deploying AI: "You want to make sure that you have an enterprise data warehouse with clean, curated data, which should be supporting all of your traditional BI and analytics workloads, before you go and start hiring a lot of data scientists and initiating a lot of generative AI projects. If organizations don't start by building strong data foundations, their data scientists will squander their time on basic data integration and cleanup."

But haven't we been here before, and often? Bond recalls that, over the last three decades of his career, the aspiration to collect and combine data into a single view of reality has been a constant one – if never fully achieved.

"If organizations don't start by building strong data foundations, their data scientists will squander their time on basic data integration and cleanup."

George Fraser, CEO, Fivetran

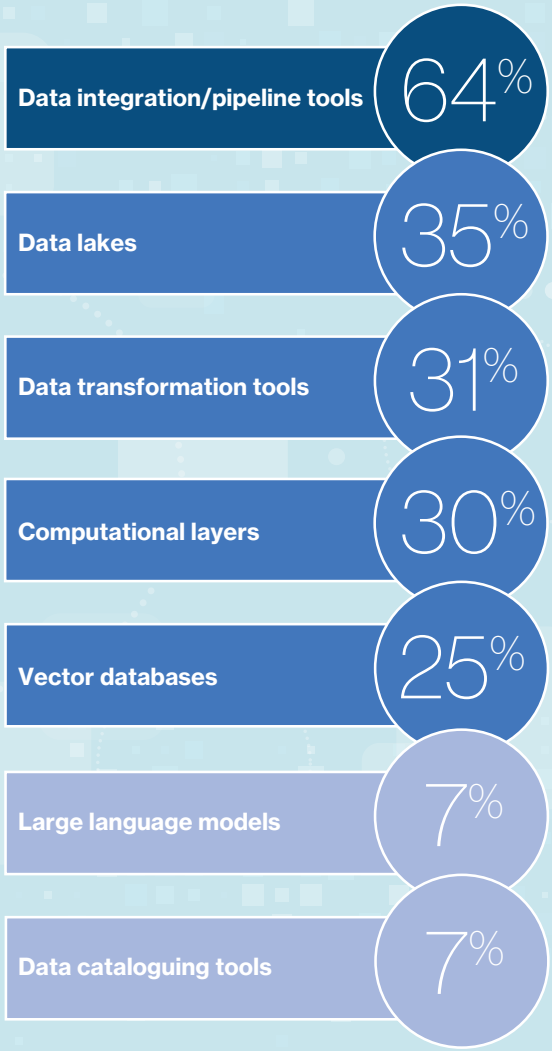
Several factors seem to have been at play. First, as Bond notes, the task of data integration and aggregation keeps getting harder. Potential data sources, file formats, and modes of use continue to multiply. Adding to the complexity are the appearance of new technologies without the old disappearing: "The technical debt out there is very real," he says.

Meanwhile, earlier work on integration does not necessarily meet the requirements for new uses. Previous business efforts around data have typically been commissioned in ad hoc fashion, to solve an immediate issue or address a specific use case, rather than with the idea of supporting a full range of present and future data needs. As such, organizations with AI ambitions are now making data foundations a top priority, investing in data tools and repositories above all (see Figure 2).

This time, however, companies are looking to use their AI-related data investments to move beyond these costly patterns of the past. Among survey respondents, 82% say that they "are prioritizing acquiring data integration and data movement solutions that will continue to work in the future, regardless of other changes to our data strategy and partners." It seems that executives now understand the value of investing in a comprehensive data foundation built on best-in-class technology that can support the organization's data and AI aspirations in the long term.

Figure 2. Top investment priorities when deploying AI

How would you prioritize the following areas of investment for an organization seeking to deploy AI? (Respondents ranking in top two.)



Source: MIT Technology Review Insights survey, 2024

03

The many challenges of data readiness

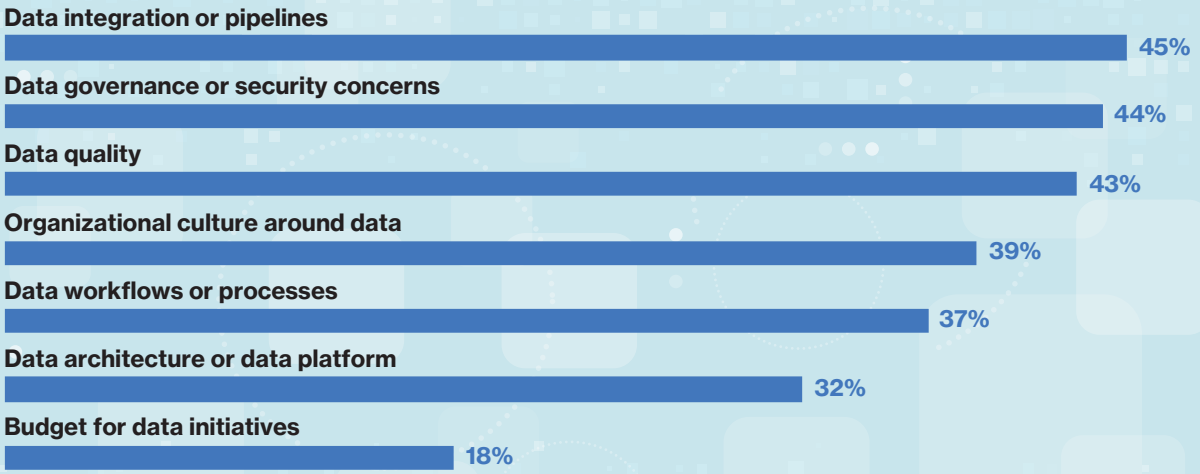
Data readiness – centralizing an organization's data and preparing it for use – is a complicated venture. The data needed for AI projects, particularly, is voluminous and sometimes rapidly changing; originating from many sources; in different degrees of cleanliness and completeness; and structured, semi-structured, or unstructured. As Bond puts it, "data is all over the place. It's very diverse, with so many different kinds, types, formats, and modes. It's also very dynamic, always moving and changing."

Hite describes what this looks like in practice. In a single week, he reports, Saks "considers a wide range of roughly a half a billion lines of data. That's how you're interacting with marketing campaigns; how you're interacting with the site; all of your post-purchase experience, like calls to the call center; as well as everything that's happening in our fulfillment center and in our broader network."

When asked about their primary data readiness challenges (see Figure 3), roughly four in 10 respondents included each of the following: data integration or pipelines (45%); data governance or security concerns (44%); data quality (43%); and organizational culture around data (39%).

Figure 3. The many difficulties in preparing data for AI

What are the primary challenges your organization has encountered in preparing its data for AI or generative AI applications?



Source: MIT Technology Review Insights survey, 2024

This report will look more deeply at the first two: data integration and data governance and security. Downplaying the importance of the others, however, would be wrong: at least one of these difficulties affects 89% of respondents' companies.

The challenge of data quality, cited by 43%, is multi-faceted. "Data needs to be clean, consistent, correct, and timely, and you have to be able to use it in the appropriate context," says Bond. "The reality is, though, that data will never be all of these things at the same time." Bad data also incurs real costs: A study by Vanson Bourne and Fivetran found that AI models built on low-quality data cost companies an average of 6% of annual revenue.[5]

Organizational culture around data may seem a surprising concern among survey respondents, given that the value of data has long been recognized. But beyond valuing data, organizations need to continually rethink value creation as technological capabilities evolve. Venkatarayalu says that Honeywell is constantly innovating and using data to add value for its customers. For example, he says, "Honeywell is one of the largest manufacturers of sensors in the world. How can we take that data source and create something that is meaningful for our customers and Honeywell?"

Hite, meanwhile, explains that using data in new ways requires "a mind shift" to see all of it as potentially important. "It's easy to put human bias on top of the types of data you feed the algorithm to produce insights," he says. "Instead, you need to let the programs tell you what data is important and how that is driving an outcome. It is not as easy as you would think." Another necessary attitudinal change, he adds, is not letting past failures blind companies to the realization that IT tools "have finally caught up to the ambitions of data scientists and data engineers."

A lack of such a cultural understanding of data impedes executives from exploring the new possibilities it can present. Indeed, those who report that culture is a primary challenge in data preparation are more than one-third more likely than others not to see the scaling of AI use as a priority and equally more likely to say that budget is a primary obstacle to data preparation for AI. This link suggests that attitudes may play as big a role in restricting spending on data readiness as might a simple lack of company resources.

A large majority of respondents, however, are committed to securing the benefits that expanded AI use is expected to bring. For them, getting data integration and governance and security right are necessary foundations.

C-suite leaders say the number-one challenge in achieving AI readiness is data integration and pipelines (45%).

The complexity of centralizing and integrating enterprise data

Survey respondents acknowledge that there's little point in trying to build AI tools before the necessary data integration foundations are in place. When asked to prioritize investments for a company focused on rolling out AI, their most common selection, by far, was data integration or data pipeline tools, with 64% of respondents including them in their top two. Data lakes came in a distant second (35%). In contrast, only 7% mentioned LLMs – the core element of generative AI platforms (see Figure 2).

As noted previously, survey respondents also say that data integration and pipelines are their primary data readiness challenge. Additionally, a large majority (77%) agree that "data integration or data movement has been a significant challenge for our organization."

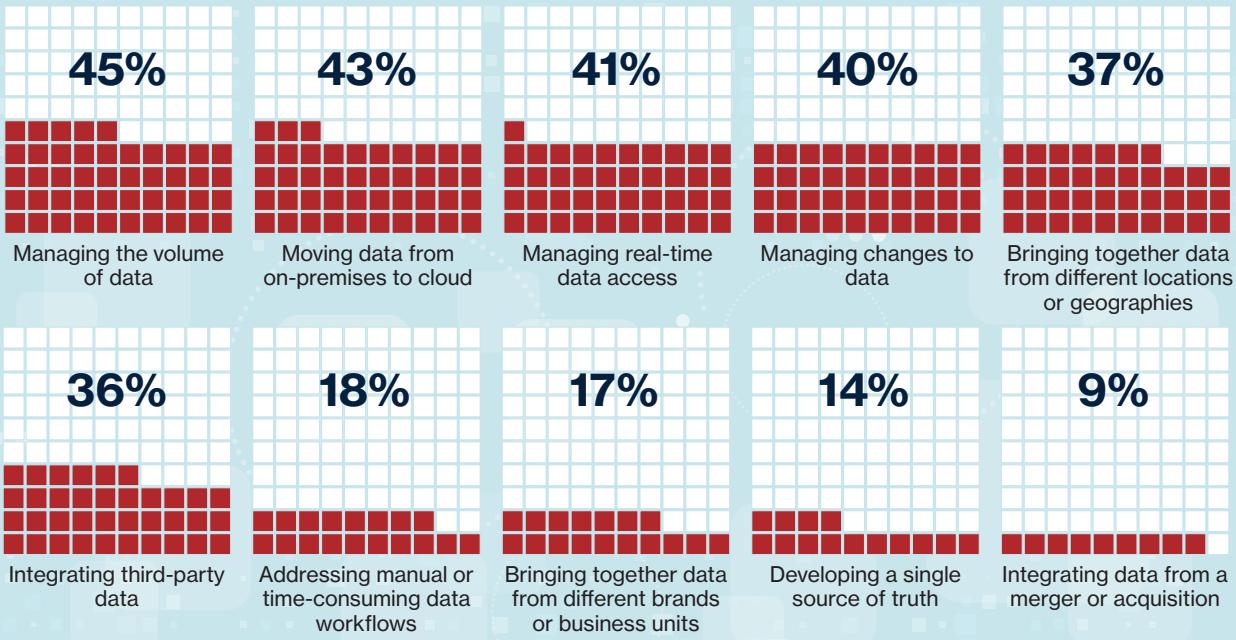
This breadth and complexity of the data required for AI explains the wide range of data integration difficulties companies face (see Figure 4).

Four – all related to bringing together immense amounts of information – are cited as primary challenges by nearly equal numbers: managing the volume of data (45%); moving data from on-premises to the cloud (43%); managing real-time data access (41%); and managing changes to data (40%).

Close behind are two items that reflect the challenges of leveraging information from outside the core of the organization: bringing together data from different locations or geographies (37%) and integrating third-party data (36%).

Figure 4. Numerous data integration challenges

What are the primary challenges your organization has encountered with data integration?



Source: MIT Technology Review Insights survey, 2024

Toward a more lasting approach

What does good look like in addressing these data challenges, especially if the aim is to establish a solid foundation for AI that remains useful even as use cases and partnerships evolve?

At a basic level, says Fraser, "data needs to be all in one place, accurate, up-to-date, and in a schema that the people at your company understand." Bond adds that a centralized approach to data storage also allows businesses "to put proper controls on it."

Hite, meanwhile, stresses that the changing economics of data storage should reshape thinking about its management. He explains that, in the past, additional data storage came at a cost. This made it important to transform information within data pipelines so that it could be stored in a consolidated form. Technological advances, however, have eliminated this marginal cost, he says. Instead, companies "are paying for utilization. The days of having to massage data to land in a shape where it's already useful are gone. You can land it in as raw a form as possible." As Saks has found, this shift brings with it important cost efficiencies (see sidebar).

Fraser notes that a unified and general-purpose approach to data platforms and tools can lead to better overall data quality. When data infrastructure and integration work is shared, he says, "the company as a whole is getting more and more proficient at working with your internal data. And that means you're building a foundation not just for the project you're doing now, but for whatever project you'll be doing in the future."

> "Our single purposes for data integration are making sure that the data from each application is complete and our data lake is as real-time as possible."
>
> **Mike Hite**, Chief Technology Officer, Saks

Saks: Faster access to data opens new possibilities for analytics and AI

The luxury retailer Saks has transformed its data integration processes to generate more and more useful business insight. Faster access to real-time data gives the data team both the tools and the time to experiment with finding business value, including in deploying AI models and applications.

The wide range of information in Saks' data lake, says chief technology officer Mike Hite, now allows the company "to pivot really quickly and easily" when doing analysis. Moreover, if the data team sees a new data source that might be interesting, it no longer requires manual pipeline building. Instead, Hite says, "we can pull that in and run some models to find out if that data's useful or not. We don't burn a ton of time." Any failures are fast and low cost, allowing for more experimentation, and enabling the top- and bottom-line benefits such trials can bring.

To reshape how it aggregates information, Hite says, Saks has greatly simplified its data integration process. Previously, this required extracting data from the application, translating it into a common format, transforming and cleaning it, and loading it into a repository. Now, the company has adopted a more modern approach that keeps everything in its native, raw form. "Our single purposes for data integration," he says, "are making sure that the data from each application is complete and our data lake is as real-time as possible." Then the company can transform the raw data sitting in its data lake "in lots of different ways that may or may not be useful, like into a KPI or sales metrics, or include it directly in an LLM or predictive AI model."

This strategy has also increased speed and efficiency and reduced costs. Where it once would have required several months to onboard data, it now takes as little as an hour, says Hite. The staff time freed up by this has allowed "a huge step change in the way that we think about and utilize data," says Hite. "We now spend the bulk of our time understanding data" rather than consolidating it – a far more valuable activity for the company.

05

Governance, compliance, and security: Underappreciated foundations

Our survey data, at first glance, shows an executive appreciation for the importance of data governance and security. Sixty percent of respondents agree that "our organization will need to rectify data governance, trust, and security issues before it can achieve its AI goals"; only 22% disagree. Similarly, when asked about the primary challenges in preparing data for AI, "data governance or security concerns" is the second most frequent answer (given by 44% of respondents), almost equal to first-place "data integration or pipelines" (45%).

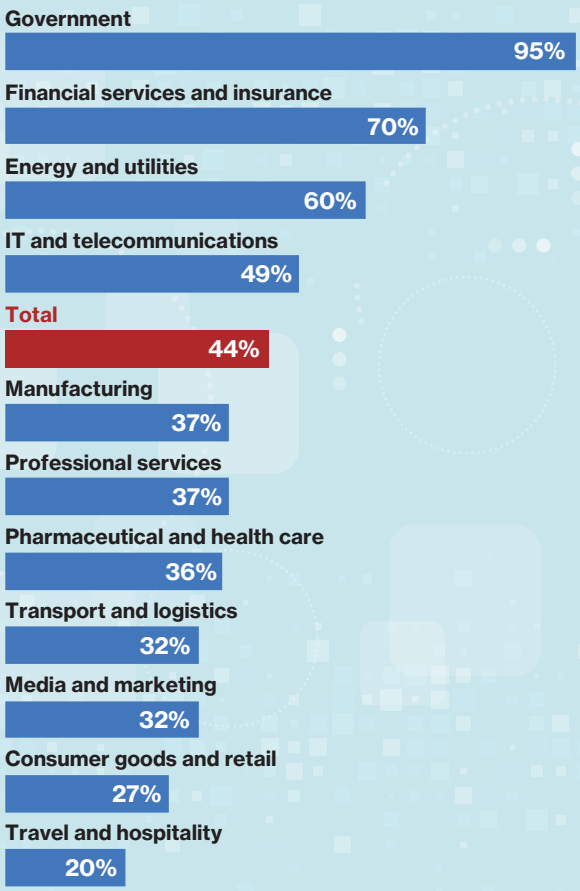
A closer look, however, shows a deep split in the responses. Concern about security and governance is highly focused among government and financial services respondents and among CDOs across industries. It is significantly less acute among the other respondents.

Respondents from highly regulated sectors, notably government and financial services, are the most concerned about data governance and security (see Figure 5). Among government respondents, 95% call governance and security a primary challenge in preparing data for AI; of those surveyed from financial services, the figure is 70%. In the rest of the survey pool, just 36% give this answer.

The other subgroup prioritizing security and governance is CDOs, of whom 74% list governance and security as a primary challenge of data preparation for AI. This is twice the rate of most of their C-suite peers: 36% of CEOs choose this answer, 38% of chief technology officers, 47% of chief information officers, and 36% of other C-level executives.

Figure 5. Concern about data governance and security varies widely by industry

Percentage of respondents who say data governance or security concerns are a primary challenge of preparing data for AI



Source: MIT Technology Review Insights survey, 2024

Hite confirms this split in interests. “When talking to boards or other leaders in a business, it’s often ‘outcome, outcome, outcome,’” he says. “But among technology executives, the first thing we always talk about as a group is governance. And that is the one thing that is absolutely not talked about enough.’”

When asked about the most important attributes of these data solutions, 100% of government and 63% of financial services respondents included built-in security and governance among the most important attributes for data solutions, as did 58% of CDOs. Among the first two groups, this is by far the most common answer; for CDOs, it comes a close second to support for cloud and on-premises solutions (61%).

Our experts suggest that those most concerned see the issue clearly even if others do not. Bond explains that most companies are using prebuilt models for generative AI, typically adding in their own data and sending further information in prompts. He warns that “organizations may have no control over someone using a piece of data in a business application and sending it to a generative AI model. These are critical concerns.” Hite adds that, while some sectors may face tighter regulation, these issues apply to any large company with customer information: “For Saks, we have a customer data governance framework in place that helps us ensure we are responsible with our customers’ data and use it to better serve our customers.”

While more attention to governance and security provisions in data solutions would presumably help in this area, the issue requires more than technological fixes. As Bond concludes, “the more you can have policies and control around data in an operational store, the better off you’ll be.” Data governance and security is essential when starting to build out new solutions incorporating AI or generative AI (see sidebar).

“When talking to boards or other leaders in a business, it’s often ‘outcome, outcome, outcome.’ But among technology executives, the first thing we always talk about as a group is governance.”

Mike Hite, Chief Technology Officer, Saks

Honeywell: Good governance enables generative AI deployment

Suresh Venkatarayalu, chief technology and innovation officer at Honeywell, explains that the company has long embedded deterministic AI within its core products, such as Honeywell Forge. For this technology, questions of governance and information security are largely settled, he says: “We control it altogether – the data, the training, the inferences.”

Now, Venkatarayalu reports that the company is focused on its first major generative AI project. By bringing together decades worth of data from disparate systems across its support business, including service tickets, product manuals, maintenance records, knowledge articles, and technical publications, Honeywell aims to train an AI model that can assist the company’s maintenance engineers, service engineers, and operators. Honeywell believes this opportunity is significant: Venkatarayalu says that he sees a potential to augment the knowledge of technicians and reduce the risk of production and operations losses.

Generative AI also brings new control and compliance challenges and opportunities. At the core of Honeywell’s generative AI strategy is a foundation of responsible AI governance. A Data and AI Steering Council made up of leaders across the organization meets monthly to drive the overall governance, development, and deployment strategy.

The maintenance project described above gives a good example of what this means in practice. Even as those involved in it are still examining fundamental questions, such as which data to include, what to build in-house, and how to measure accuracy and benefits, the governance is largely in place. “At this juncture,” Venkatarayalu reports, the team, along with its generative AI partner, has “put enough guardrails and controls in place so that we are comfortable that we are not revealing the overall IP, including proprietary data and domain knowledge.”

06 Conclusion

The heightened competition around the deployment of generative AI has again brought corporate data into focus. This time, though, executives understand the necessity of creating a lasting data foundation that can serve both current and future technologies. Our research has identified three key principles for doing so.

Data foundations must precede AI deployment

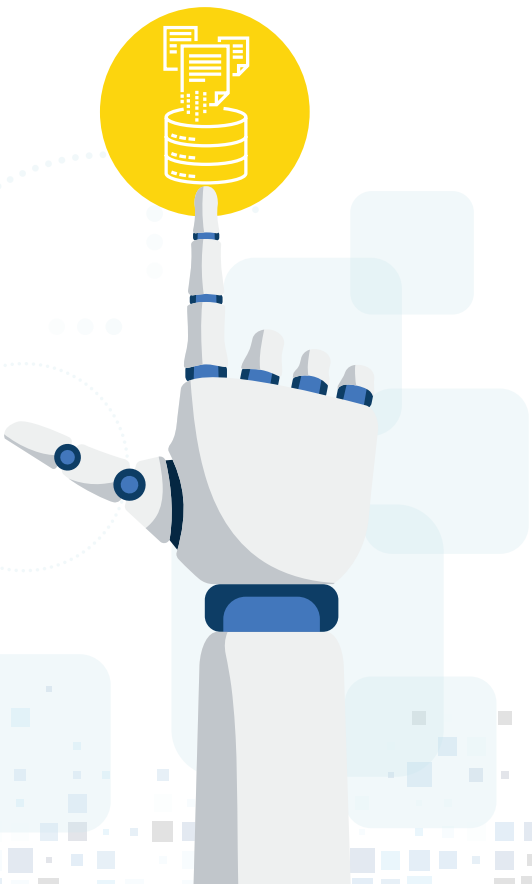
Most organizations will need to address the technical debt in their data infrastructure before they can even begin developing impactful AI applications. Respondents to our survey advise organizations seeking to implement AI to invest first in data integration, pipelines, and data lakes.

Data integration is the fundamental activity of data readiness

The most impactful uses of AI will rely on businesses' ability to utilize their unique and proprietary data. Survey respondents report that bringing together disparate stores of data is the biggest challenge of preparing their data for AI – making data integration capabilities a key differentiator for producing business value.

Data governance and security should be addressed from day one

CDOs and survey respondents from regulated industries already see what the rest of us will come to understand: data governance and security are fundamental to AI projects. Data workflows and tools that consider governance and security first – rather than addressing these as an afterthought – are necessary for the most responsible and compliant use of these technologies, but they are also most likely to result in successful AI deployments at scale.



About MIT Technology Review Insights

MIT Technology Review Insights is the custom publishing division of *MIT Technology Review,* the world's longest-running technology magazine, backed by the world's foremost technology institution – producing live events and research on the leading technology and business challenges of the day. Insights conducts qualitative and quantitative research and analysis in the US and abroad and publishes a wide variety of content, including articles, reports, infographics, videos, and podcasts. And through its growing MIT Technology Review Global Insights Panel, Insights has unparalleled access to senior-level executives, innovators, and entrepreneurs worldwide for surveys and in-depth interviews.

About Fivetran

Fivetran, the global leader in data movement, helps customers use their data to power everything from AI applications and ML models, to predictive analytics and operational workloads. The Fivetran platform reliably and securely centralizes data from hundreds of SaaS applications and databases into any cloud destination – whether deployed on-premises, in the cloud, or in a hybrid environment. Thousands of global brands, including Autodesk, Condé Nast, JetBlue, and Morgan Stanley, trust Fivetran to move their most valuable data assets to fuel analytics, drive operational efficiencies, and power innovation.

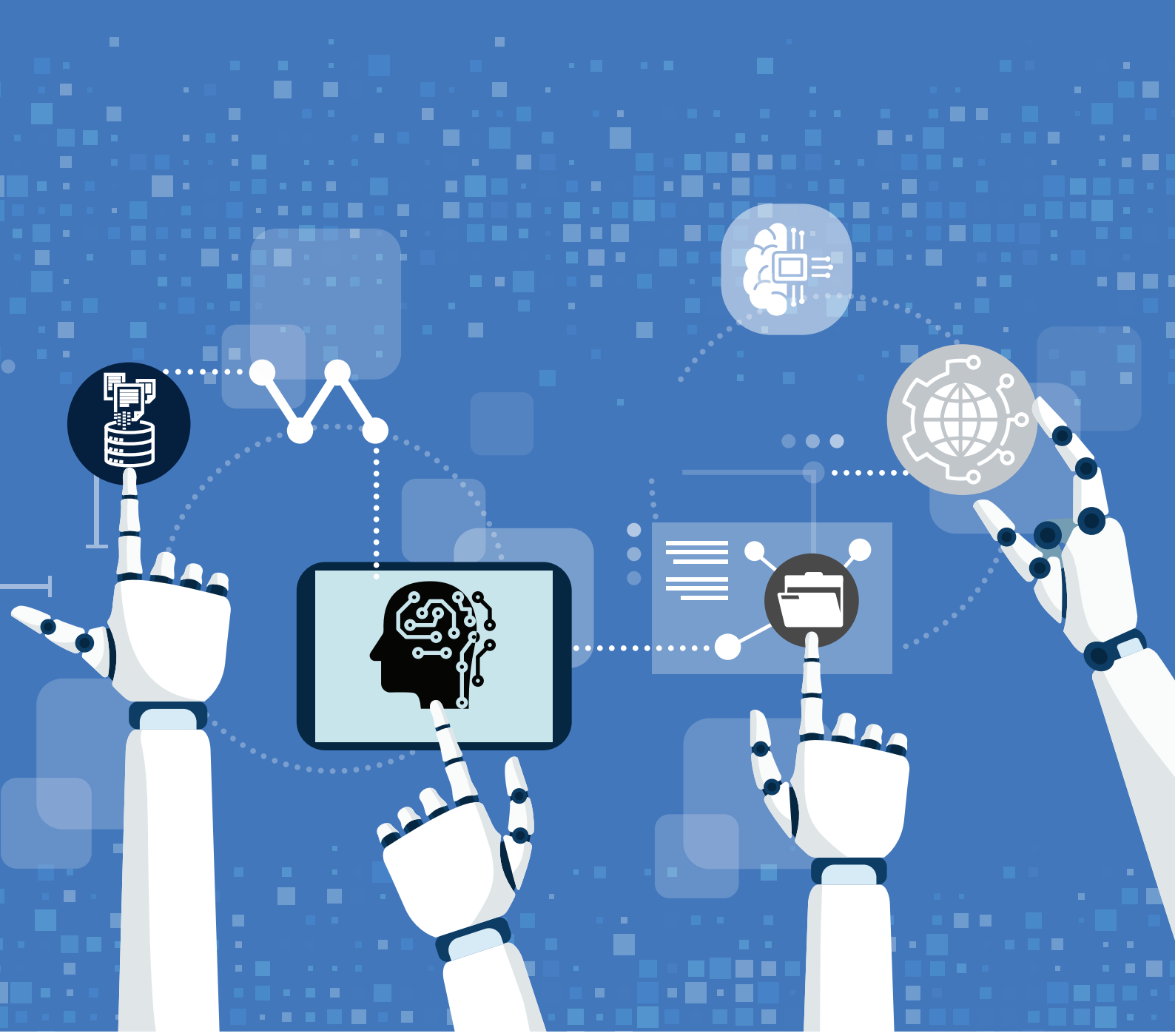
For more info, visit **fivetran.com**.

Endtnotes

1. McKinsey & Company, "The economic potential of generative AI: The next productivity frontier," June 2023, https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#introduction.
2. MIT Technology Review, "Generative AI: Differentiating disruptors from the disrupted," 2024, https://www.technologyreview.com/2024/02/29/1089152/generative-ai-differentiating-disruptors-from-the-disrupted/.
3. Boston Consulting Group, *From Potential to Profit with GenAI*, January 2024, https://www.bcg.com/publications/2024/from-potential-to-profit-with-genai.
4. Gartner, "Gartner Poll Finds 55% of Organizations are in Piloting or Production Mode with Generative AI." press release, October 2023, https://www.gartner.com/en/newsroom/press-releases/2023-10-03-gartner-poll-finds-55-percent-of-organizations-are-in-piloting-or-production-mode-with-generative-ai.
5. Mark Van de Wiel, "New AI survey: Poor data quality leads to $406 million in losses," Fivetran, March 20, 2024, https://www.fivetran.com/blog/new-ai-survey-poor-data-quality-leads-to-406-million-in-losses.

Illustrations

Illustrations assembled by Tim Huxford with elements from Shutterstock and Adobe Stock.



MIT Technology Review Insights

www.technologyreview.com

insights@technologyreview.com